

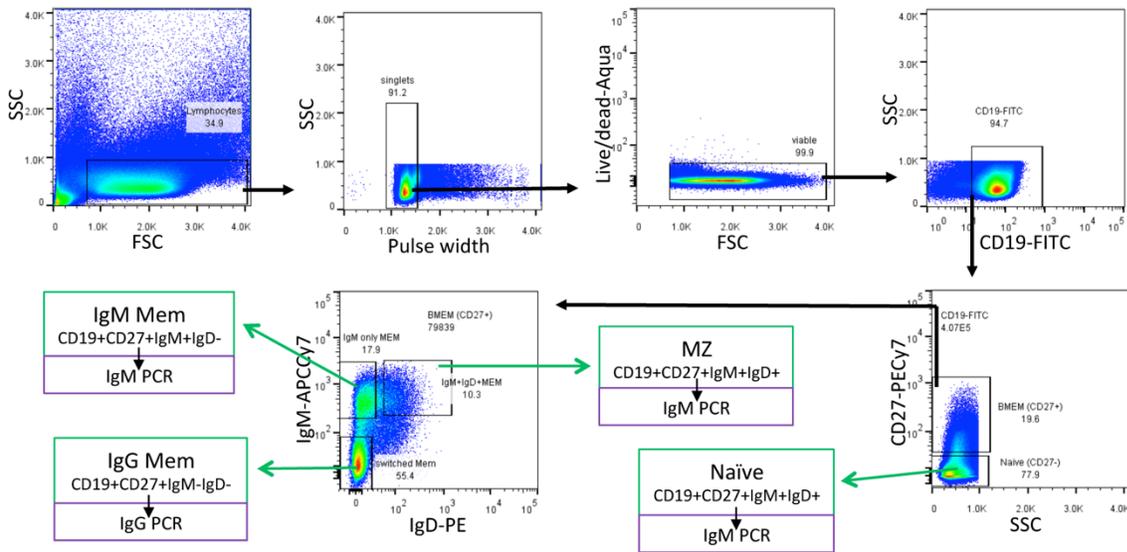
Supporting Information

Table S1. Overview of samples used for sequencing, and the number of sequences obtained from each sample. Visit 1 is day 0, Visit 2 is day 7, Visit 3 is day 28, and Visit 4 is day 35.

Group	Participant	Visit	Cell subset	Isotype	Cell number	Total sequences	Sequences after filtering	Clusters
Conjugate	1	V1	Naïve	IgM	140000	147455	104389	50971
			MZ	IgM	19000	140810	110489	16124
			IgM Mem	IgM	14000	150452	118839	7281
			IgG Mem	IgG	44000	116925	66685	7382
		V2	PC DR+	IgM	4152	149225	111649	1105
			PC DR-	IgG	3134	130565	65223	1566
				IgM	3134	142544	112549	2277
	V4	PC DR+	IgG	191	107066	66595	647	
	2	V1	Naïve	IgM	70000	143386	101837	37903
			MZ	IgM	2643	144168	104228	5761
			IgM Mem	IgM	10000	147730	119636	8907
			IgG Mem	IgG	12000	118918	66299	4051
		V2	PC DR+	IgG	2369	109273	65857	1371
				IgM	2369	133201	104043	891
			PC DR-	IgG	6462	100022	69469	1410
	V4	PC DR-	IgM	6462	158416	126141	1814	
	V4	PC DR+	IgG	110	116052	76475	436	
	3	V3	Naïve	IgM	22000	147142	108238	47265
			MZ	IgM	17000	137404	108214	12887
			IgM Mem	IgM	19000	143281	110264	9673
			IgG Mem	IgG	30000	118486	74469	4406
		V2	PC DR+	IgG	34000	131585	73360	2786
				IgM	34000	120966	73869	1208
			PC DR-	IgG	1400	149622	93959	1167
V4	PC DR-	IgM	1400	140271	113928	1211		
V4	PC DR+	IgG	1222	113482	73270	2139		
4	V3	Naïve	IgM	100000	150584	108669	52442	
		MZ	IgM	11000	164406	120101	11669	
		IgM Mem	IgM	9015	154232	118549	6973	
		IgG Mem	IgG	30000	111747	75530	4378	
	V2	PC DR+	IgG	2536	122207	80731	1710	
			IgM	2536	160339	126600	1378	
		PC DR-	IgG	905	127309	74675	1209	
V4	PC DR-	IgM	905	145834	117478	862		
V4	PC DR+	IgG	1623	120930	76302	2055		
Polysaccharide	5	V1	Naïve	IgM	97000	149982	103711	47578
			MZ	IgM	1224	147823	118588	3861
			IgM Mem	IgM	1888	139335	106526	3365
			IgG Mem	IgG	2940	122018	79345	1308
		V2	PC DR+	IgG	2860	119590	78235	853
				IgM	2860	145868	119941	2571
			PC DR-	IgG	10000	106711	79434	893
	V4	PC DR-	IgM	10000	106686	81546	1046	
	V4	PC DR+	IgG	538	126470	75958	982	
	6	V1	Naïve	IgM	100000	139241	96086	43565
MZ			IgM	2787	119250	93460	5220	

		IgM Mem	IgM	2783	143934	113632	5343
		IgG Mem	IgG	100000	114550	73642	2714
	V2	PC DR+	IgG	2852	113866	78498	922
			IgM	2852	142966	111665	1625
	PC DR-	IgG	IgM	9081	121421	73844	1285
			IgM	9081	154623	115372	1457
	V4	PC DR+	IgG	657	120998	77809	1280
7	V1	Naïve	IgM	64000	134352	106141	28237
		MZ	IgM	2612	154606	120293	6197
		IgM Mem	IgM	4846	143613	114253	3096
		IgG Mem	IgG	15000	108576	67579	3481
	V2	PC DR+	IgG	3025	105639	70821	821
			IgM	3025	139101	112616	1091
		PC DR-	IgG	737	111782	81456	913
			IgM	737	137767	107742	1056
V4	PC DR+	IgG	1765	120886	71007	1443	
8	V3	Naïve	IgM	100000	158575	110940	42659
		MZ	IgM	2157	145540	112070	4576
		IgM Mem	IgM	8294	135451	104517	7374
		IgG Mem	IgG	5766	120279	76277	3070
	V2	PC DR+	IgG	1984	119126	88336	1724
			IgM	1984	142845	109075	2076
		PC DR-	IgG	221	116541	77417	601
			IgM	221	149833	123045	813
V4	PC DR+	IgG	7491	100654	67928	1562	
9	V3	Naïve	IgM	12000	150115	111326	19245
		MZ	IgM	2350	162209	126365	4570
		IgM Mem	IgM	2801	148589	117019	3792
		IgG Mem	IgG	4598	116989	83930	2077
	V2	PC DR+	IgG	50000	113195	63911	1712
			IgM	50000	157507	124651	2078
		PC DR-	IgG	11000	117897	78888	1565
			IgM	11000	153629	116653	1518
V4	PC DR+	IgG	1514	117743	86175	1662	

Visit 1 and Visit 3 cell sorting



Visit 2 and Visit 4 cell sorting

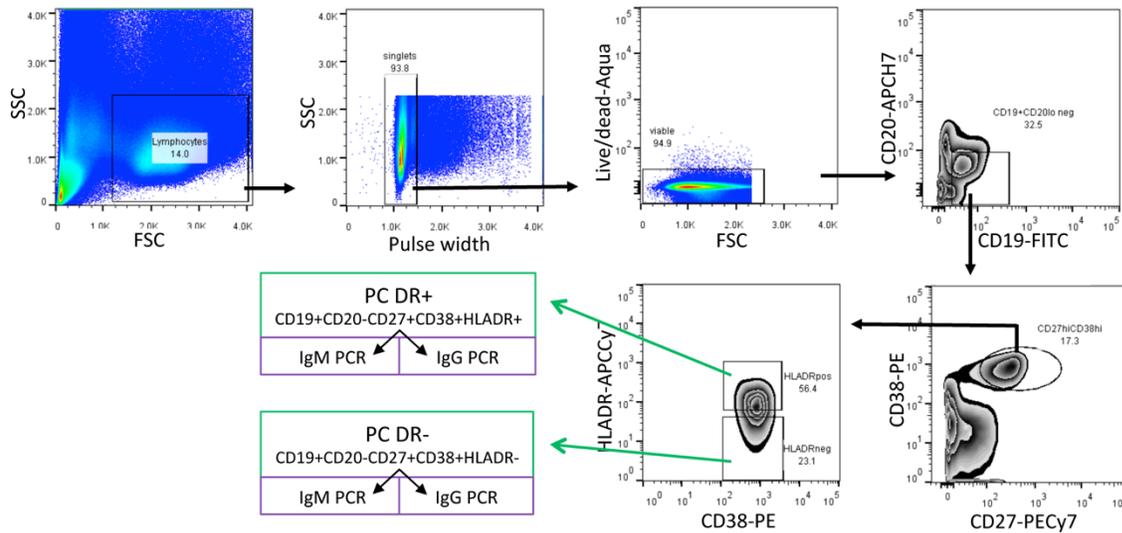


Fig. S1. Representative FACS plots showing the gating strategy for sorting the B cell subsets at the different visits. On the vaccine administration visits (Visit 1 and Visit 3), naïve, marginal zone (MZ), IgM memory and IgG memory B cell subsets were sorted based on their expression of CD19, CD27, IgM and IgD surface markers. IgM-specific antibody transcripts were PCR amplified from the naïve, MZ and IgM memory subsets, and IgG-specific antibody transcripts from the IgG memory subset for deep sequencing. On the visits seven days post vaccination (Visit 2 and Visit 4), HLA-DR+ (PC DR+) and HLA-DR- (PC DR-) PCs were sorted based on their expression of CD19, CD20, CD27, CD38 and HLA-DR surface markers. Both IgM and IgG antibody transcripts were independently amplified from both PC subsets for deep sequencing.

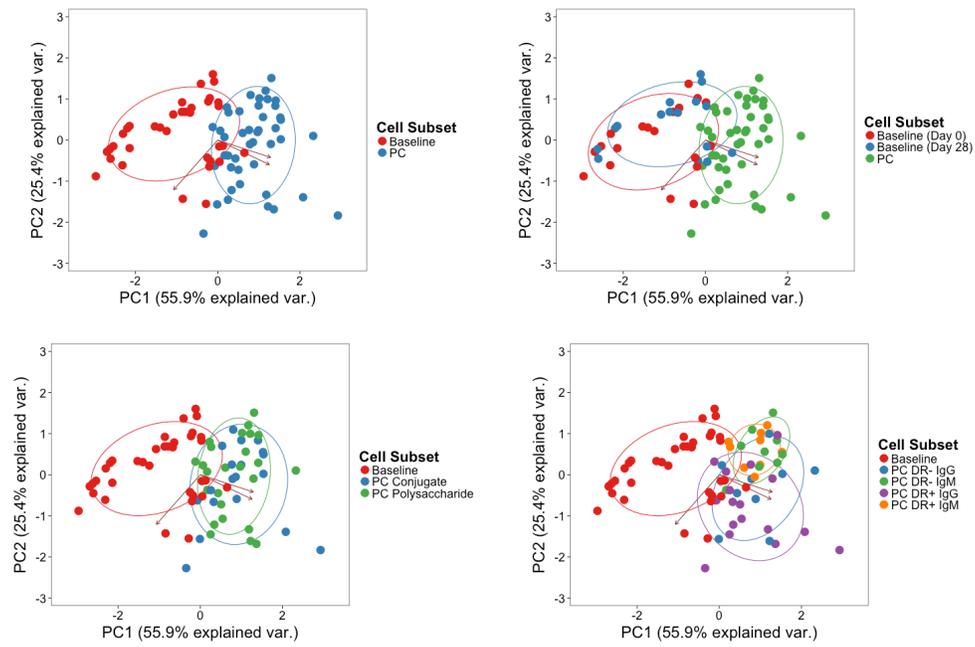


Fig. S2. Principal component analysis including all cell subsets, utilizing the clonality, V gene mutation, and CDR3 length variables. Different plots show the different cell subsets colored according to different properties.

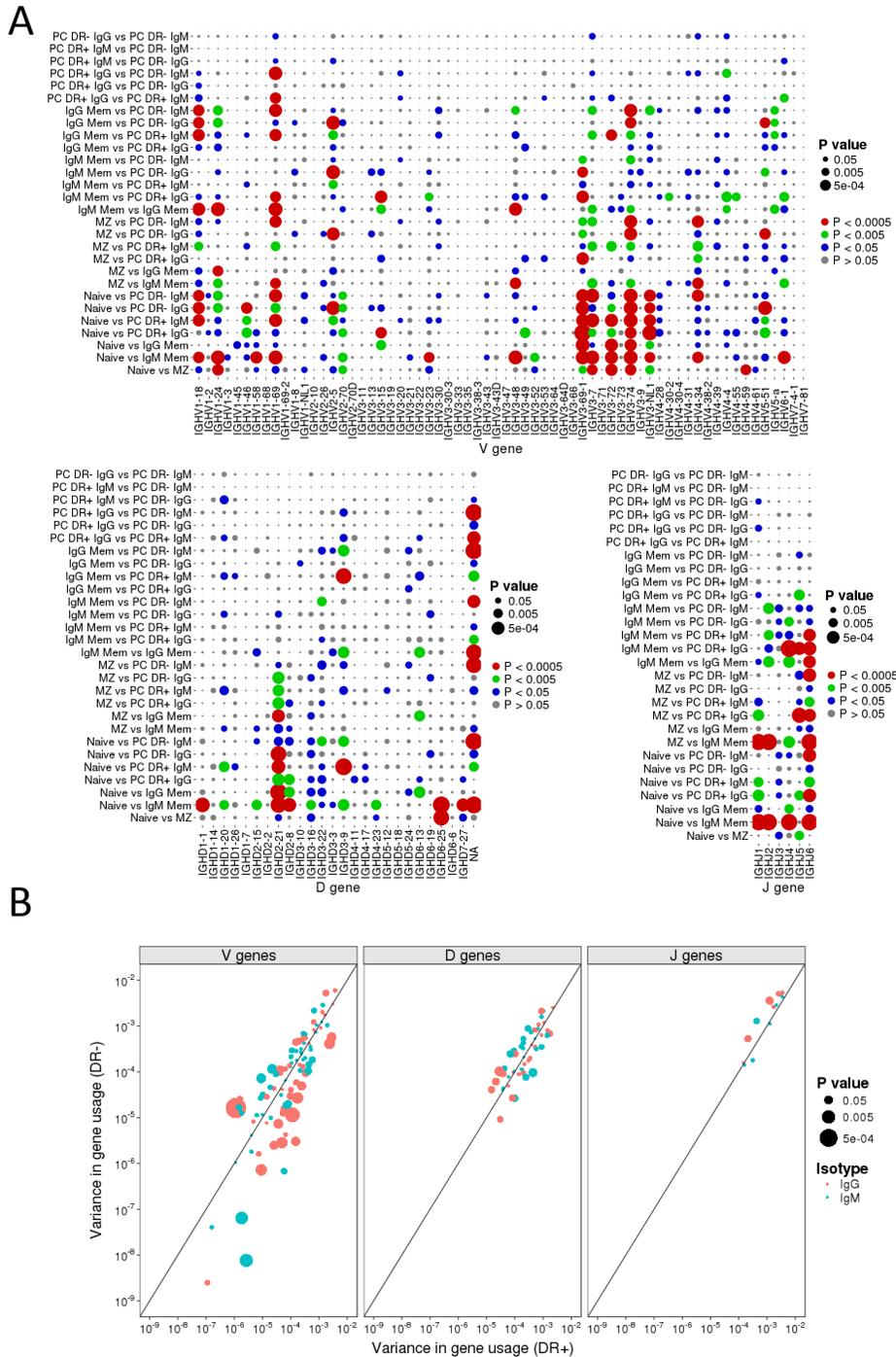


Fig. S3. Differences VDJ usage between the different cell subsets. A) Differences in proportions of V, D, and J gene usage between B cell subsets. Size and color of bubble indicates level of significance. Comparisons performed using the pairwise Mann-Whitney U test with Bonferroni correction for multiple comparisons. B) Comparison of variance in V, D and J gene usage between the different participants for HLA-DR+ and HLA-DR- PCs. Each bubble represents a V, D or J gene segment, with its location determined by the variance in the use of that gene segment in the two PC populations. Thus, for bubbles above the line of equality, variance is greater for that gene segment usage between the different individuals for the HLA-DR- compared to the HLA-DR+ PCs. The size of the bubble represents the p-value after an F-test for equality of variances in the two PC populations.

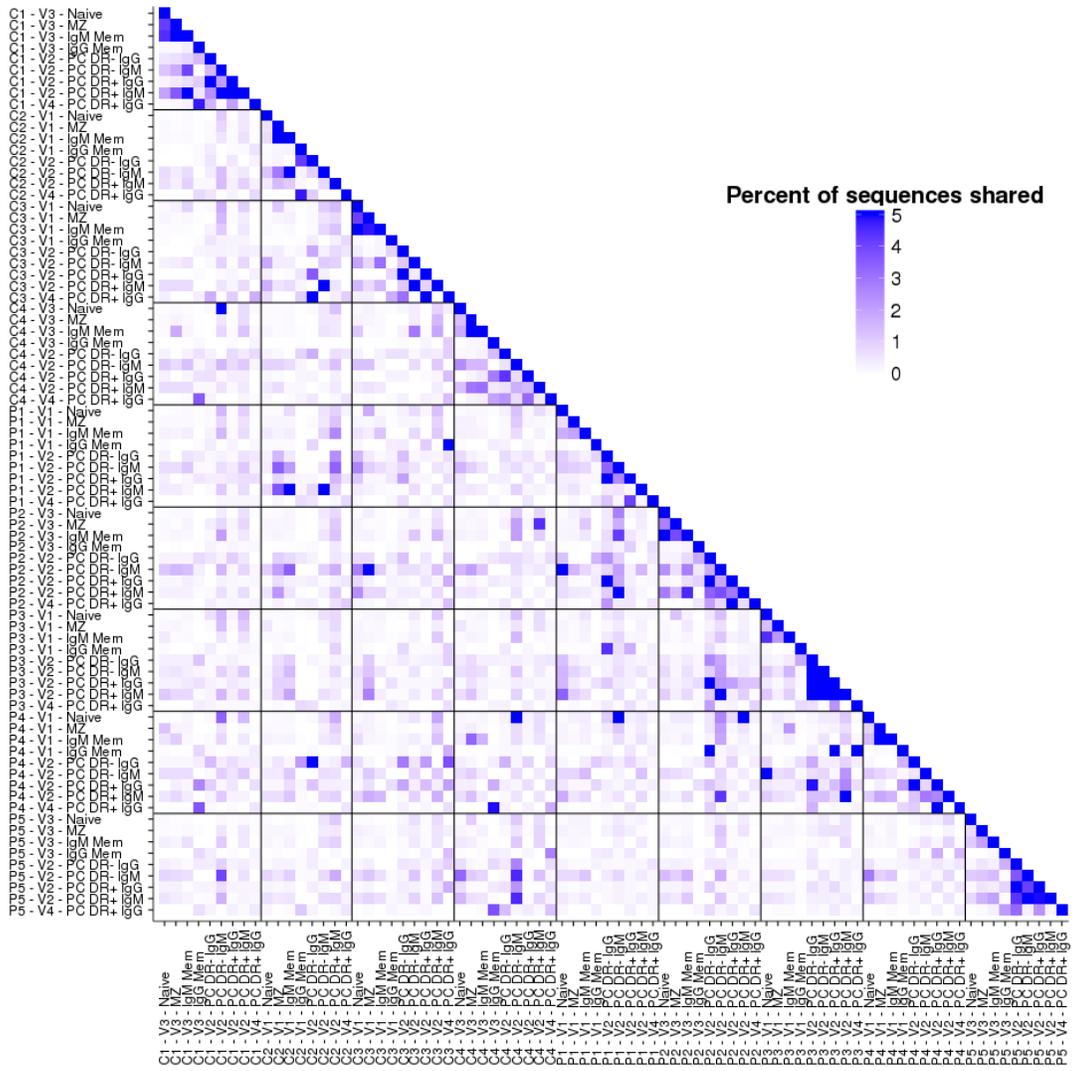


Fig. S4. Percent of sequences shared between all different samples $[(A \cap B) / \min(A, B)]$ sequenced. White squares indicate low overlap between samples, and blue squares indicate high overlap. For sequences to be defined as shared they had to have an exact CDR3 AA sequence match. “P” and “C” notations in the axis labels denote participants from the Polysaccharide and Conjugate groups respectively.

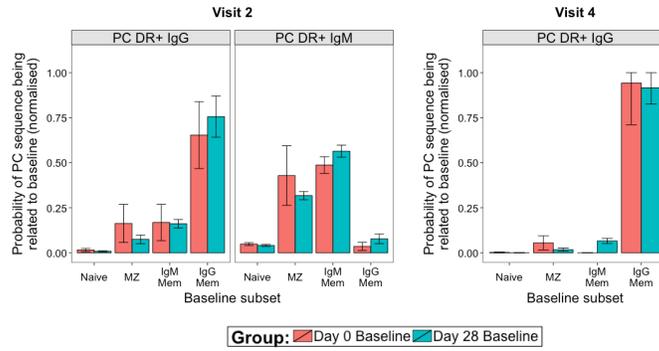


Fig. S5. Probability of shared PC sequences being shared with each baseline cell subset for HLA-DR+ IgG and IgM PCs at visit 2, after the first vaccine, and for HLA-DR+ IgG PCs at visit 4, after the second vaccine. For sequences to be defined as shared they had to have an exact CDR3 AA sequence match, and use the same V and J genes. Groups correspond to participants where baseline subsets were isolated at day 0 (N = 5) vs day 28 (N = 4). Error bars indicate \pm SEM.

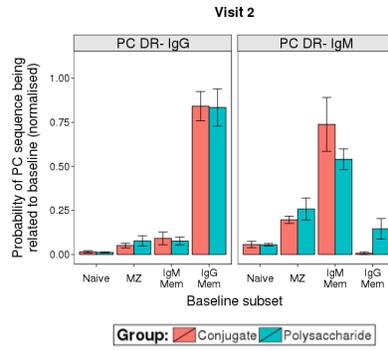


Fig. S6. Probability of shared PC sequences being shared with each baseline cell subset for HLA-DR- IgG and IgM PCs at visit 2, after the first vaccine. For sequences to be defined as shared they had to have an exact CDR3 AA sequence match, and use the same V and J genes. Error bars indicate \pm SEM.

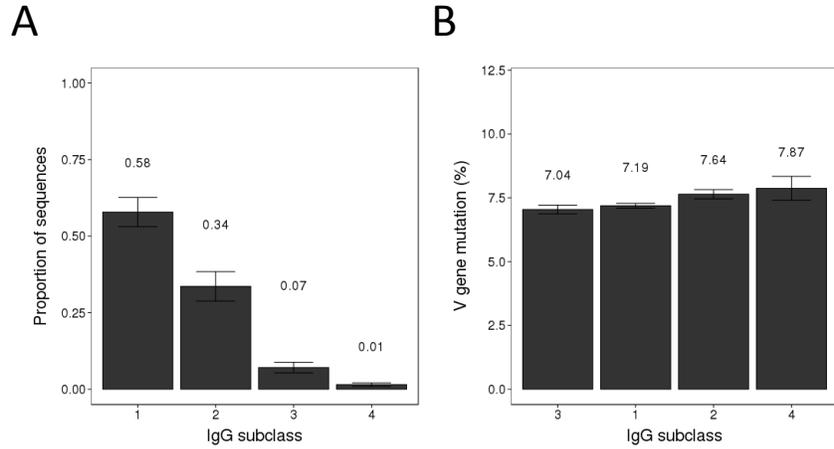


Fig. S7. IgG isotype subclass and mutation levels using the IgG memory population as a baseline control. A) Mean proportion of sequences of each IgG subclass, ordered in sequence of decreasing proportions (1→2→3→4). B) Mean percent of mutated nucleotides in the V gene of sequences belonging to each IgG subclass, ordered in sequence of increasing mutation (3→1→2→4). For A & B, mean values are noted above each bar. Error bars indicate \pm SEM.